



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Participant Subjectivity and Involvement as a Basis for Discourse Segmentation

Citation for published version:

Niekrasz, J & Moore, J 2009, Participant Subjectivity and Involvement as a Basis for Discourse Segmentation. in *Proceedings of the SIGDIAL 2009 Conference*. pp. 54-62.
<<http://portal.acm.org/citation.cfm?id=1708384>>

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Proceedings of the SIGDIAL 2009 Conference

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Participant Subjectivity and Involvement as a Basis for Discourse Segmentation

John Niekrasz and Johanna Moore

Human Communication Research Centre

School of Informatics

University of Edinburgh

{jniekras,jmoore}@inf.ed.ac.uk

Abstract

We propose a framework for analyzing episodic conversational activities in terms of expressed relationships between the participants and utterance content. We test the hypothesis that linguistic features which express such properties, e.g. tense, aspect, and person deixis, are a useful basis for automatic intentional discourse segmentation. We present a novel algorithm and test our hypothesis on a set of intentionally segmented conversational monologues. Our algorithm performs better than a simple baseline and as well as or better than well-known lexical-semantic segmentation methods.

1 Introduction

This paper concerns the analysis of conversations in terms of communicative activities. Examples of the kinds of activities we are interested in include relating a personal experience, making a group decision, committing to future action, and giving instructions. The reason we are interested in these kinds of events is that they are part of participants' common-sense notion of the goals and accomplishments of a dialogue. They are part of participants' subjective experience of what happened and show up in summaries of conversations such as meeting minutes. We therefore consider them an ideal target for the practical, common-sense description of conversations.

Activities like these commonly occur as cohesive *episodes* of multiple turns within a conversation (Korolija, 1998). They represent an intermediate level of dialogue structure – greater than a single

speech act but still small enough to have a potentially well-defined singular purpose. They have a temporal granularity of anywhere from a few seconds to several minutes.

Ultimately, it would be useful to use descriptions of such activities in automatic summarization technologies for conversational genres. This would provide an activity-oriented summary describing what 'happened' that would complement one based on information content or what the conversation was 'about'. Part of our research goal is thus to identify a set of discourse features for segmenting, classifying, and describing conversations in this way.

1.1 Participant subjectivity and involvement

The approach we take to this problem is founded upon two basic ideas. The first is that the activities we are interested in represent a coarse level of the *intentional structure* of dialogue (Grosz and Sidner, 1986). In other words, each activity is unified by a common purpose that is shared between the participants. This suggests there may be linguistic properties which are shared amongst the utterances of a given activity episode.

The second idea concerns the properties which distinguish different activity types. We propose that activity types may be usefully distinguished according to two complex properties of utterances, both of which concern relationships between the participants and the utterance: *participant subjectivity* and *participant involvement*. Participant subjectivity concerns attitudinal and perspectival relationships *toward* the dialogue content. This includes properties such as whether the utterance expresses

the private mental state of the speaker, or the participants’ temporal relationship to a described event. Participant involvement concerns the roles participants play *within* the dialogue content, e.g., as the agent of a described event.

1.2 Intentional segmentation

The hypothesis we test in this paper is that the linguistic phenomena which express participant-relational properties may be used as an effective means of intentional discourse segmentation. This is based on the idea that if adjacent discourse segments have different activity types, then they are distinguishable by participant-relational features. If we can reliably extract such features, then this would allow segmentation of the dialogue accordingly.

We test our hypothesis by constructing an algorithm and examining its performance on an existing set of intentionally segmented conversational monologues (i.e., one person speaks while another listens) (Passonneau and Litman, 1997, henceforth P&L). While our long term goal is to apply our techniques to multi-party conversations (and to a somewhat coarser-grained analysis), using this dataset is a stepping-stone toward that end which allows us to compare our results with existing intentional segmentation algorithms.

An example dialogue extract from the dataset is shown in Dialogue 1. Two horizontal lines indicate a segment boundary which was identified by at least 3 of 7 annotators. A single horizontal line indicates a segment boundary which was identified by 2 or fewer annotators. In the example, there are three basic types of discourse activity distinguishable according to the properties of participant subjectivity and participant involvement. The segments beginning at 22.1 and 26.2 share the use of the historical present tense – a type of participant subjectivity – in a narrative activity type. Utterances 24.1 and 25.1, on the other hand, are about the prior perceptions of the speaker, a type of participant involvement in a past event. The segment beginning at 24.2 is a type of generic description activity, exhibiting its own distinct configuration of participant relational features, such as the generic *you* and present tense.

We structure the rest of the paper as follows. First, we begin by describing related and supporting theoretical work. This is followed by a test of our main

PearStories-09 (Chafe, 1980)

21.2	okay.
22.1	Meanwhile,
22.2	there are three little boys,
22.3	up on the road a little bit,
22.4	and they see this little accident.
23.1	And u-h they come over,
23.2	and they help him,
23.3	and you know,
23.4	help him pick up the pears and everything.
24.1	A-nd the one thing that struck me about the- three
	little boys that were there,
24.2	is that one had ay uh I don't know what you call
	them,
24.3	but it's a paddle,
24.4	and a ball-,
24.5	is attached to the paddle,
24.6	and you know you bounce it?
25.1	And that sound was really prominent.
26.1	Well anyway,
26.2	so- u-m tsk all the pears are picked up,
26.3	and he's on his way again,

Dialogue 1: An example dialogue extract showing intentional segment boundaries.

hypothesis. We then follow this with a similar experiment which contextualizes our work both theoretically and in practical terms with respect to the most commonly studied segmentation task: topic segmentation. We finish with a general discussion of the implications of our experiments.

2 Background and Related Work

The influential work of Grosz and Sidner (1986) provides a helpful starting point for understanding our approach. Their theory suggests that intentions (which equate to the goals and purposes of a dialogue) are a foundation for the structure of discourse. The individual discourse purposes that emerge in a dialogue relate directly to the natural aggregation of utterances into discourse segments. The attentional state of the dialogue, which contains salient objects and relations and allows for the efficient generation and interpretation of utterances, is then dependent upon this interrelated intentional and linguistic structure in the emerging dialogue.

Grosz and Sidner’s theory suggests that attentional state is parasitic upon the underlying intentional structure. This implication has informed many approaches which relate referring expressions (an attentional phenomenon) to discourse structure.

One example is Centering theory (Grosz et al., 1995), which concerns the relationship of referring expressions to discourse coherence. Another is P&L, who demonstrated that co-reference and inferred relations between noun phrases are a useful basis for automatic intentional segmentation.

Our approach expands on this by highlighting the fact that objects that are in focus within the attentional state have an important quality which may be exploited: they are focused upon *by* the participants from particular *points of view*. In addition, the objects may in fact *be* the participants themselves. We would expect the linguistic features which express such relationships (e.g., aspect, subjectivity, modality, and person deixis) to therefore correlate with intentional structure, and to do so in a way which is important to participants' subjective experience of the dialogue.

This approach is supported by a theory put forth by Chafe (1994), who describes how speakers can express ideas from alternative perspectives. For example, a subject who is recounting the events of a movie of a man picking pears might say "the man was picking pears", "the man picks some pears", or "you see a man picking pears." Each variant is an expression of the same idea but reflects a different perspective toward, or manner of participation in, the described event. The linguistic variation one sees in this example is in the properties of tense and aspect in the main clause (and in the last variant, a perspectival superordinate clause which uses the generic *you*). We have observed that discourse coheres in these perspectival terms, with shifts of perspective usually occurring at intentional boundaries.

Wiebe (1994; 1995) has investigated a phenomenon closely related to this: point-of-view and subjectivity in fictional narrative. She notes that paragraph-level blocks of text often share a common *objective* or *subjective* context. That is, sentences may or may not be conveyed from the point-of-view of individuals, e.g., the author or the characters within the narrative. Sentences continue, resume, or initiate such contexts, and she develops automatic methods for determining when the contexts shift and whose point-of-view is being taken. Her algorithm provides a detailed method for analyzing written fiction, but has not been developed for conversational or non-narrative genres.

Smith's (2003) analysis of texts, however, draws a more general set of connections between the content of sentences and types of discourse segments. She does this by analyzing texts at the level of short passages and determines a non-exhaustive list of five basic "discourse modes" occurring at that level: narrative, description, report, information, and argument. The mode of a passage is determined by the type of situations described in the text (e.g., event, state, general stative, etc.) and the temporal progression of the situations in the discourse. Situation types are in turn organized according to the perspectival properties of aspect and temporal location. A narrative passage, for example, relates principally specific events and states, with dynamic temporal advancement of narrative time between sentences. On the other hand, an information passage relates primarily general statives with atemporal progression.

3 Automatic Segmentation Experiment

The analysis described in the previous sections suggests that participant-relational features correlate with the intentional structure of discourse. In this section we describe an experiment which tests the hypothesis that a small set of such features, i.e., tense, aspect, and first- and second-person pronouns, are a useful basis for intentional segmentation.

3.1 Data

Our experiment uses the same dataset as P&L, a corpus of 20 spoken narrative monologues known as the Pear Stories (Chafe, 1980). Chafe asked subjects to view a silent movie and then summarize it for a second person. Their speech was then manually transcribed and segmented into prosodic phrases. This resulted in a mean 100 phrases per narrative and a mean 6.7 words per phrase. P&L later had each narrative segmented by seven annotators according to an informal definition of communicative intention. Each prosodic phrase boundary was a possible discourse segment boundary. Using Cochran's Q test, they concluded that an appropriate gold standard could be produced by using the set of boundaries assigned by at least three of the seven annotators. This is the gold standard we use in this paper. It assigns a boundary at a mean 16.9% ($\sigma = 4.5\%$) of the

possible boundary sites in each narrative. The result is a mean discourse segment length of 5.9 prosodic phrases, ($\sigma = 1.4$ across the means of each narrative).

3.2 Algorithm

The basic idea behind our algorithm is to distinguish utterances according to the type of activity in which they occur. To do this, we identify a set of utterance properties relating to participant subjectivity and participant involvement, according to which activity types may be distinguished. We then develop a routine for automatically extracting the linguistic features which indicate such properties. Finally, the dialogue is segmented at locations of high discontinuity in that feature space. The algorithm works in four phases: pre-processing, feature extraction, similarity measurement, and boundary assignment.

3.2.1 Pre-processing

For pre-processing, disfluencies are removed by deleting repeated strings of words and incomplete words. The transcript is then parsed (Klein and Manning, 2002), and a collection of typed grammatical dependencies are generated (de Marneffe et al., 2006). The TTT2 chunker (Grover and Tobin, 2006) is then used to perform tense and aspect tagging.

3.2.2 Feature extraction

Feature extraction is the most important and novel part of our algorithm. Each prosodic phrase (the corpus uses prosodic phrases as sentence-like units, see Data section) is assigned values for five binary features. The extracted features correspond to a set of utterance properties which were identified manually through corpus analysis. The first four relate directly to individual activity types and are therefore mutually exclusive properties.

first-person participation [1P] – helps to distinguish meta-discussion between the speaker and hearer (e.g., “Did I tell you that?”)

generic second-person [2P-GEN] – helps to distinguish narration told from the perspective of a generic participant (e.g., “You see a man picking pears”)

third-person stative/progressive [3P-STAT] – helps to distinguish narrative activities related

to “setting the scene” (e.g., “[There is a man | a man is] picking pears”)

third-person event [3P-EVENT] – helps to distinguish event-driven third-person narrative activities (e.g. “The man drops the pears”)

past/non-past [PAST] – helps to distinguish narrative activities by temporal orientation (e.g. “There man drops the pears” vs. “The man dropped the pears”)

Feature extraction works by identifying the linguistic elements that indicate each utterance property. First, prosodic phrases containing a first- or second-person pronoun in grammatical subject or object relation to any clause are identified (common fillers like *you know*, *I think*, and *I don’t know* are ignored). Of the identified phrases, those with first-person pronouns are marked for 1P, while the others are marked for 2P-GEN. For the remaining prosodic phrases, those with a matrix clause are identified. Of those identified, if either its head verb is *be* or *have*, it is tagged by TTT2 as having progressive aspect, or the prosodic phrase contains an existential *there*, then it is marked for 3P-STAT. The others are marked for 3P-EVENT. Finally, if the matrix clause was tagged as past tense, the phrase is marked for PAST. In cases where no participant-relational features are identified (e.g., no matrix clause, no pronouns), the prosodic phrase is assigned the same features as the preceding one, effectively marking a continuation of the current activity type.

3.2.3 Similarity measurement

Similarity measurement is calculated according to the cosine similarity $\cos(v_i, c_i)$ between the feature vector v_i of each prosodic phrase i and a weighted sum c_i of the feature vectors in the preceding context. The algorithm requires a parameter l to be set for the desired mean segment length. This determines the window $w = \text{floor}(l/2)$ of preceding utterances to be used. The weighted sum representing the preceding context is computed as $c_i = \sum_{j=1}^w ((1+w-j)/w) v_{i-j}$, which gives increasingly greater weight to the more recent phrase.

3.2.4 Boundary assignment

In the final step, the algorithm assigns boundaries where the similarity score is lowest, namely prior to

prosodic phrases where cos is less than the first $1/l$ quantile for that discourse.

3.3 Experimental Method and Evaluation

Our experiment compares the performance of our novel algorithm (which we call NM09) with a naive baseline and a well-known alternative method – P&L’s co-reference based NP algorithm. To our knowledge, P&L is the only existing publication describing algorithms designed specifically for intentional segmentation of dialogue. Their NP algorithm exploits annotations of direct and inferred relations between noun phrases in adjacent units. Inspired by Centering theory (Grosz et al., 1995), these annotations are used in a computational account of discourse focus to measure coherence. Although adding pause-based features improved results slightly, the NP method was the clear winner amongst those using a single feature type and produced very good results.

The NP algorithm requires co-reference annotations as input, so to create a fully-automatic version (NP-AUTO) we have employed a state-of-the-art co-reference resolution system (Poesio and Kabadjov, 2004) to generate the required input. We also include results based on P&L’s original human co-reference annotations (NP-HUMAN).

For reference, we include a baseline that randomly assigns boundaries at the same mean frequency as the gold-standard annotations, i.e., a sequence drawn from the Bernoulli distribution with success probability $p = 0.169$ (this probability determines the value of the target segment length parameter l in our own algorithm). As a top-line reference, we calculate the mean of the seven annotators’ scores with respect to the three-annotator gold standard.

For evaluation we employ two types of measure. On one hand, we use $P(k)$ (Beeferman et al., 1999) as an error measure designed to accommodate near-miss boundary assignments. It is useful because it estimates the probability that two randomly drawn points will be assigned incorrectly to either the same or different segments. On the other hand, we use Cohen’s Kappa (κ) to evaluate the precise placement of boundaries such that each potential boundary site is considered a binary classification. While κ is typically used to evaluate inter-annotator agree-

Table 1: Mean results for the 18 test narratives.

	$P(k)$	κ	F_1	Rec.	Prec.
Human	.21	.58	.65	.64	.69
NP-HUMAN	.35	.38	.40	.52	.46
NM09	.44	.11	.24	.23	.28
NP-AUTO	.52	.03	.27	.71	.17
Random	.50	.00	.15	.14	.17

ment, it is a useful measure of classification accuracy in our experiment for two reasons. First, it accounts for the strong class bias in our data. Second, it allows a direct and intuitive comparison with our inter-annotator top-line reference. We also provide results for the commonly-used IR measures F_1 , recall, and precision. These are useful for comparing with previous results in the literature and provide a more widely-understood measure of the accuracy of the results. Precision and recall are also helpful in revealing the effects of any classification bias the algorithms may have.

The results are calculated for 18 of the 20 narratives, as manual feature development involved the use of two randomly selected narratives as development data. The one exception is NP-HUMAN, which is evaluated on the 10 narratives for which there are manual co-reference annotations.

3.4 Results

The mean results for the 18 narratives, calculated in comparison to the three-annotator gold standard, are shown in Table 2. NP-HUMAN and NM09 are both superior to the random baseline for all measures ($p \leq 0.05$). NP-AUTO, however, is only superior in terms of recall and F_1 ($p \leq 0.05$).

3.5 Discussion

The results indicate that the simple set of features we have chosen can be used for intentional segmentation. While the results are not near human performance, it is encouraging that such a simple set of easily extractable features achieves results that are 19% (κ), 24% ($P(k)$), and 18% (F_1) of human performance, relative to the random baseline.

The other notable result is the very high recall score of NP-AUTO, which helps to produce a respectable F_1 score. However, a low κ reveals that

when accounting for class bias, this system is actually not far from the performance of a high recall random classifier.

Error analysis showed that the reason for the problems with NP-AUTO was the lack of reference chains produced by the automatic co-reference system. While the system seems to have performed well for direct co-reference, it did not do well with bridging reference. Inferred relations were an important part of the reference chains produced by P&L, and it is now clear that these play a significant role in the performance of the NP algorithm. Our algorithm is not dependent on this difficult processing problem, which typically requires world knowledge in the form of training on large datasets or the use of large lexical resources.

4 Topic vs. Intentional Segmentation

It is important to place our experiment on intentional segmentation in context with the most commonly studied automatic segmentation task: topic-based segmentation. While the two tasks are distinct, the literature has drawn connections between them which can at times be confusing. In this section, we attempt to clarify those connections by pointing out some of their differences and similarities. We also conduct an experiment comparing our algorithm to well-known topic-segmentation algorithms and discuss the results.

4.1 Automatic segmentation in the literature

One of the most widely-cited discourse segmentation algorithms is TextTiling (Hearst, 1997). Designed to segment texts into multi-paragraph subtopics, it works by operationalizing the notion of lexical cohesion (Halliday and Hasan, 1976). TextTiling and related algorithms exploit the collocation of semantically related lexemes to measure coherence. Recent improvements to this method include the use of alternative lexical similarity metrics like LSA (Choi et al., 2001) and alternative segmentation methods like the minimum cut model (Malioutov and Barzilay, 2006) and ranking and clustering (Choi, 2000). Recently, Bayesian approaches which model topics as a lexical generative process have been employed (Purver et al., 2006; Eisenstein and Barzilay, 2008). What these algo-

rithms all share is a focus on the semantic content of the discourse.

Passonneau and Litman (1997) is another of the most widely-cited articles on discourse segmentation. Their overall approach combines an investigation of prosodic features, cue words, and entity reference. As described above, their approach to using entity reference is motivated by Centering theory (Grosz et al., 1995) and the hypothesis that intentional structure is exhibited in the attentional relationships between discourse referents.

Hearst and P&L try to achieve different goals, but their tasks are nonetheless related. One might reasonably hypothesize, for example, that either lexical similarity or co-reference could be useful to either type of segmentation on the grounds that the two phenomena are clearly related. However, there are also clear differences of intent between the two studies. While there is an obvious difference in the dataset (written expository text vs. spoken narrative monologue), the annotation instructions reflect the difference most clearly. Hearst instructed naive annotators to mark paragraph boundaries “where the *topics* seem to change,” whereas P&L asked naive annotators to mark prosodic phrases where the speaker had begun a new *communicative task*.

The results indicate that there is a difference in granularity between the two tasks, with intentional segmentation relating to finer-grained structure. Hearst’s segments have a mean of about 200 words to P&L’s 40. Also, two hierarchical topic segmentations of meetings (Hsueh, 2008; ?) have averages above 400 words for the smallest level of segment.

To our knowledge, P&L is the only existing study of automatic intention-based segmentation. However, their work has been frequently cited as a study of topic-oriented segmentation, e.g., (Galley et al., 2003; Eisenstein and Barzilay, 2008). Also, recent research in conversational genres (Galley et al., 2003; Hsueh and Moore, 2007) analyze events like discussing an agenda or giving a presentation, which resemble more intentional categories. Interestingly, these algorithms demonstrate the benefit of including non-lexical, non-semantic features. The results imply that further analysis is needed to understand the links between different types of coherence and different types of segmentation.

Table 2: Results comprising our method to topic-oriented segmentation methods.

NP-auto	$P(k)$	κ	F_1	Rec.	Prec.
Human	.21	.58	.65	.64	.69
NM09	.44	.11	.24	.24	.28
C99	.44	.08	.22	.20	.24
TEXTTILING	.41	.05	.18	.16	.21
Random	.50	.00	.15	.14	.17

4.2 Experiment 2

We have extended the above experiment to compare the results of our novel algorithm with existing topic segmentation methods. We employ Choi’s implementations of C99 (Choi, 2000) and TEXTTILING (Hearst, 1997) as examples of well-known topic-oriented methods. While we acknowledge that there are newer algorithms which improve upon this work, these were selected for being well studied and easy to apply out-of-the-box. Our method and evaluation is the same as in the previous experiment.

The mean results for the 18 narratives are shown in Table 2, with the human and baseline score reproduced from the previous table. All three automatic algorithms are superior to the random baseline in terms of $P(k)$, κ , and F_1 ($p \leq 0.05$). The only statistically significant difference ($p \leq 0.05$) between the three automatic methods is between NM09 and TEXTTILING in terms of F_1 . The observed difference between NM09 and TEXTTILING in terms of κ is only moderately significant ($p \leq 0.08$). The observed differences between NM09 and C99 are minimally significant ($p \leq 0.24$).

4.3 Discussion

The comparable performance achieved by our simple perspective-based approach in comparison to lexical-semantic approaches suggests two main points. First, it validates our novel approach in practical applied terms. It shows that perspective-oriented features, being simple to extract and applicable to a variety of genres, are potentially very useful for automatic discourse segmentation systems.

Second, the results show that the teasing apart of topic-oriented and intentional structure may be quite difficult. Studies of coherence at the level of short

passages or episodes (Korolija, 1998) suggest that coherence is established through a complex interaction of topical, intentional, and other contextual factors. In this experiment, the major portion of the dialogues are oriented toward the basic narrative activity which is the premise of the Pear Stories dataset. This means that there are many times when the activity type does not change at intentional boundaries. At other times, the activity type changes but neither the topic nor the set of referents is significantly changed. The different types of algorithms we have tried (i.e., topical, referential, and perspectival) seem to be operating on somewhat orthogonal bases, though it is difficult to say quantitatively how this relates to the types of “communicative task” transitions occurring at the boundaries. In a sense, we have proposed an algorithm for performing “activity type cohesion” which mimics the methods of lexical cohesion but is based upon a different dimension of the discourse. The results indicate that these are both related to intentional structure.

5 General Discussion and Future Work

Future work in intentional segmentation is needed. Our ultimate goal is to extend this work to more conversational domains (e.g., multi-party planning meetings) and to define the richer set of perspectives and related deictic features that would be needed for them. For example, we hypothesize that the different uses of second-person pronouns in conversations (Gupta et al., 2007) are likely to reflect alternative activity types. Our feature set and extraction methods will therefore need to be further developed to capture this complexity.

The other question we would like to address is the relationship between various types of coherence (e.g., topical, referential, perspectival, etc.) and different types (and levels) of discourse structure. Our current approach uses a feature space that is orthogonal to most existing segmentation methods. This has allowed us to gain a deeper understanding of the relationship between certain linguistic features and the underlying intentional structure, but more work is needed.

In terms of practical motivations, we also plan to address the open question of how to effectively combine our feature set with other feature sets which

have also been demonstrated to contribute to discourse structuring and segmentation.

References

- Doug Beeferman, Adam Berger, and John D. Lafferty. 1999. Statistical models for text segmentation. *Machine Learning*, 34(1-3):177–210.
- Wallace L. Chafe, editor. 1980. *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*, volume 3 of *Advances in Discourse Processes*. Ablex, Norwood, NJ.
- Wallace L. Chafe. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. University of Chicago Press, Chicago.
- Freddy Y. Y. Choi, Peter Wiemer-Hastings, and Johanna Moore. 2001. Latent semantic analysis for text segmentation. In *Proc. EMNLP*, pages 109–117.
- Freddy Y. Y. Choi. 2000. Advances in domain independent linear text segmentation. In *Proc. NAACL*, pages 26–33.
- Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Proc. LREC*, pages 562–569.
- Jacob Eisenstein and Regina Barzilay. 2008. Bayesian unsupervised topic segmentation. In *Proc. EMNLP*, pages 334–343.
- Michel Galley, Kathleen McKeown, Eric Fosler-Lussier, and Hongyan Jing. 2003. Discourse segmentation of multi-party conversation. In *Proc. ACL*, pages 562–569.
- Barbara J. Grosz and Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.
- Barbara J. Grosz, Aravind Joshi, and Scott Weinstein. 1995. Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.
- Claire Grover and Richard Tobin. 2006. Rule-based chunking and reusability. In *Proc. LREC*.
- Alexander Gruenstein, John Niekrasz, and Matthew Purver. 2008. Meeting structure annotation: Annotations collected with a general purpose toolkit. In L. Dybkjaer and W. Minker, editors, *Recent Trends in Discourse and Dialogue*, pages 247–274.
- Surabhi Gupta, John Niekrasz, Matthew Purver, and Daniel Jurafsky. 2007. Resolving “you” in multi-party dialog. In *Proc. SIGdial*, pages 227–230.
- M. A. K. Halliday and Ruqayia Hasan. 1976. *Cohesion in English*. Longman, New York.
- Marti Hearst. 1997. TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1):33–64.
- Pei-Yun Hsueh and Johanna D. Moore. 2007. Combining multiple knowledge sources for dialogue segmentation in multimedia archives. In *Proc. ACL*, pages 1016–1023.
- Pei-Yun Hsueh. 2008. *Meeting Decision Detection: Multimodal Information Fusion for Multi-Party Dialogue Understanding*. Ph.D. thesis, School of Informatics, University of Edinburgh.
- Dan Klein and Christopher D. Manning. 2002. Fast exact inference with a factored model for natural language parsing. In *NIPS 15*.
- Natascha Korolija. 1998. *Episodes in talk: Constructing coherence in multiparty conversation*. Ph.D. thesis, Linköping University, The Tema Institute, Department of Communications Studies.
- Igor Malioutov and Regina Barzilay. 2006. Minimum cut model for spoken lecture segmentation. In *Proc. COLING-ACL*, pages 25–32.
- Rebecca J. Passonneau and Diane J. Litman. 1997. Discourse segmentation by human and automated means. *Computational Linguistics*, 23(1):103–139.
- Massimo Poesio and Mijail A. Kabadjov. 2004. A general-purpose, off-the-shelf anaphora resolution module: Implementation and preliminary evaluation. In *Proc. LREC*.
- Matthew Purver, Konrad Körding, Thomas Griffiths, and Joshua Tenenbaum. 2006. Unsupervised topic modelling for multi-party spoken discourse. In *Proc. COLING-ACL*, pages 17–24.
- Carlota S. Smith. 2003. *Modes of Discourse*. Cambridge University Press, Cambridge.
- Janyce M. Wiebe. 1994. Tracking point of view in narrative. *Computational Linguistics*, 20(2):233–287.
- Janyce M. Wiebe. 1995. References in narrative text. In Judy Duchan, Gail Bruder, and Lynne Hewitt, editors, *Deixis in Narrative: A Cognitive Science Perspective*, pages 263–286.